

2020

The effect of the length of a questionnaire on the accuracy estimations of the ability and the psychometric properties of the item and scale in the light of item response theory

Abedalhameed Alqudah

Ministry of Education and Higher Education, Qatar, sarsak2002@yahoo.com

Nedal Al-Shraifin

Yarmouk University,

Follow this and additional works at: https://digitalcommons.aaru.edu.jo/anujsr_b

Recommended Citation

Alqudah, Abedalhameed and Al-Shraifin, Nedal (2020) "The effect of the length of a questionnaire on the accuracy estimations of the ability and the psychometric properties of the item and scale in the light of item response theory," *An-Najah University Journal for Research - B (Humanities)*: Vol. 34 : Iss. 6 , Article 2.

Available at: https://digitalcommons.aaru.edu.jo/anujsr_b/vol34/iss6/2

This Article is brought to you for free and open access by Arab Journals Platform. It has been accepted for inclusion in An-Najah University Journal for Research - B (Humanities) by an authorized editor. The journal is hosted on [Digital Commons](#), an Elsevier platform. For more information, please contact rakan@aarj.edu.jo, marah@aarj.edu.jo, u.murad@aarj.edu.jo.

أثر طول الاستبانة على دقة تقديرات القدرة والخصائص السيكومترية للفقرة والمقياس في ضوء نظرية الاستجابة للفقرة†

The effect of the length of a questionnaire on the accuracy estimations of the ability and the psychometric properties of the item and scale in the light of item response theory

عبد الحميد القضاة*، ونضال الشرايفين**

Abedalhameed Alqudah & Nedal Al-Shraifin

*طالب دكتوراة، جامعة اليرموك، حالياً: وزارة التعليم والتعليم العالي، قطر. **قسم علم النفس التربوي، كلية التربية، جامعة اليرموك، إربد، الأردن

* PHD. student, Yarmouk University, currently: Ministry of Education and Higher Education, Qatar. ** Department of Educational Psychology, College of Education, Yarmouk University, Irbid, Jordan

* الباحث المراسل: sarsak2002@yahoo.com

تاريخ التسليم: (2018/8/16)، تاريخ القبول: (2018/11/12)

ملخص

هدفت الدراسة للتعرف إلى أثر طول الاستبانة على دقة تقدير معالم القدرة للأفراد، ومعالم الفقرات والمقياس وفق نموذج الاستجابة المتدرجة (GRM). فتم اختيار عينة الدراسة بالطريقة المُتبصرة؛ وذلك بتوزيع نماذج الاستبانة على المعلمين، حيث بلغ عدد أفراد العينة (1425) معلماً ومعلمة، منهم (1038) معلماً و(387) معلمة؛ موزعين إلى (492) معلماً ومعلمة لنموذج الاستبانة (10 فقرات) و(497) معلماً ومعلمة لنموذج الاستبانة (30 فقرة) و(436) معلماً ومعلمة لنموذج الاستبانة (60 فقرة)، كما تم استخدام استبانة للرؤى الوظيفي، وأظهرت النتائج وجود فروق دالة إحصائية عند مستوى الدلالة ($\alpha=0.05$) بين الأوساط الحسابية للأخطاء المعيارية في تقدير قدرات الأفراد يُعزى لطول الاستبانة؛ وأنَّ الفروق الجوهرية بين الأوساط الحسابية للأخطاء المعيارية وفقاً لطول الاستبانة كانت على الترتيب: لصالح طول الاستبانة (60) مقارنةً بطول الاستبانة (10) ثم (30)، ولصالح طول الاستبانة (30) مقارنةً بطول الاستبانة (10). وأظهرت النتائج أيضاً وجود فروق دالة

† هذا البحث مستل من أطروحة الدكتوراة للطالب عبد الحميد القضاة بعنوان "أثر طول الاستبانة على دقة تقدير معالم القدرة والخصائص السيكومترية للفقرة والمقياس في ضوء نظرية الاستجابة للفقرة" والتي تم مناقشتها في جامعة اليرموك بتاريخ 2018\12\27م..

إحصائيًا بين الأوساط الحسابية للأخطاء المعيارية في تقدير معالم التمييز لل فقرات يُعزى لطول الاستبانة؛ وأنَّ الفروق الجوهرية بين الأوساط الحسابية للأخطاء المعيارية كانت لصالح طول الاستبانة (30) مقارنةً بطول الاستبانة (10)، ثم لصالح طول الاستبانة (60) مقارنةً بطول الاستبانة (10). وأظهرت النتائج عدم وجود فروق دالة إحصائيًا بين الأوساط الحسابية للأخطاء المعيارية في تقدير معالم صعوبة الفقرات يُعزى لمعالم عتبة الصعوبة وطول الاستبانة، والتفاعل بينهما، وفي ضوء نتائج الدراسة يوصي الباحثان على مطوري الاستبانات كأداة لبحوثهم مراعاة طول الاستبانة للحصول على دقة في تقديرات القدرة للأفراد عينة الدراسة. والأفضل أن تكون من ذات (30) فقرة فأكثر.

الكلمات المفتاحية: طول الاستبانة، دقة تقديرات القدرة، نظرية استجابة الفقرة، نموذج الاستجابة المتدرجة (GRM).

Abstract

The present study aimed at identifying the effect of the length of the questionnaire on the Accuracy of Ability Estimation of the parameters of examinees and The parameters of Items and scale in the light of IRT, the sample of the study was selected in a available sample by distributing the questionnaire samples to the teachers. Amounted (1425) male and female teachers, were (1038) male and (387) female distributed to (492) teachers of the questionnaire (10 items) And (497) teachers of the questionnaire (30 items) and (436) teachers of the questionnaire model (60 items), in order to achieve the objectives of the study; Abdel-Jawad & Metwally (1993) questionnaire job satisfaction of teachers was used, The results showed statistically significant differences at the level of ($\alpha = 0.05$) between the standard error average in the estimation of the abilities of individual related to the length of the questionnaire; the results showed differences between the standard error averages in favor Length of questionnaire (60) compared to (length of questionnaire (10) and (30)), then for (30) compared to the length of the questionnaire (10). The findings also indicate that there were statistically significant differences at the level of significance ($\alpha = 0.05$) between the averages of the standard error in estimating the discrimination parameter for the item related to the length of the questionnaire; a differences between the average of the standard error in favor for the length (30) compared to the length of the questionnaire (10), and in favor of the length of the questionnaire (60) compared to the length of the questionnaire (10). Lastly the results indicate there were no statistically

significant differences between the averages of the standard error in estimating the difficulty parameter of the item due to the difficulty parameter and length of the questionnaire, and to the interaction of the difficulty parameter with the length of the questionnaire, and to the interaction of the difficulty parameter with the length of the questionnaire, in the light of the results of study the researcher recommended questionnaire developers as a tool for their research taking into account the length of the item to achieve accuracy in the speed estimates of the sample elements. It is better to be from the same paragraph (30) or more.

Keywords: Length of Questionnaire, Accuracy of Ability Estimates of Item, Item Response Theory, GRM Model.

المقدمة

تُعدّ المقاييس التربوية من أهم الأسس التي تُسهم في تحقيق أهداف البحث المرجوة، وتحظى المقاييس التربوية أيضاً باهتمام المتخصصين في مجال القياس النفسي والتربوي نظراً لاعتماد الكثير من القرارات على نتائج هذه المقاييس، حيث إنها تُقدم معلومات لأطراف ومجالات متعددة، ونظراً لأهمية القرارات التي تقدمها المقاييس التربوية والنفسية فإنه لا بد من أن تتسق هذه المقاييس مع الأهداف التي تُعد من أجلها، وأن تتمتع بقيم ثبات مرتفعة، وخاصة في المقاييس النفسية التي تتأثر بتأثير الشخصية ومن محوص إلى آخر، فإن انتقاء الفقرات يحتاج إلى دقة عالية لتخدم الغرض من بناء هذه المقاييس النفسية (عدس، 2003).

وتعد الاستبانة من أشهر المقاييس النفسية التي تستخدم على شكل واسع في البحوث التربوية والنفسية، إذ لاقت الاستبانة رواجاً كبيراً بين أوساط الباحثين والتربويين باستخدامها كأداة قياس رئيسة لجمع البيانات في بحوثهم؛ نظراً لسهولة جمع البيانات من خلالها من أماكن متباعدة ولعدد كبير من المفحوصين في آن واحد، فيتم إعداد استبانة تتضمن مجموعة من الفقرات تهدف لقياس سمة أو سلوك ما، وتستخدم الاستبانة بشكل رئيس في مجال الدراسات التي تهدف إلى استكشاف حقائق عن الممارسات الحالية، واستطلاعات الرأي العام، وميول الأفراد، وسمات الشخصية (النل وقحل، 2007). إلا أن الاستبانة كأداة قياس يشوبها بعض المآخذ منها ملل المستجيب في الاستجابة عليها خاصة في حال زيادة عدد فقراتها، التقدير العشوائي من قبل المستجيبين، عدا عن الوقت والجهد الذي يبذله الباحث في جمع المعلومات من خلالها، وهذا بدوره ربما يقود إلى ضعف في دقة النتائج التي يُؤمل الحصول عليها عند تطبيق الاستبانة؛ وللوصول إلى استبانة مثالية تخدم الهدف من بنائها، فقد عنيت النظرية الحديثة للقياس (نظرية استجابة الفقرة) (Item Response Theory) بالطرق والإجراءات التي تقود للوصول إلى هذه الأداة، والتي يمكن الوثوق بقراراتها عند الحكم على الأفراد من خلال تقديم مؤشرات إضافية عن نظيرتها الكلاسيكية (Tay, Meade & Smith, 2015). وتفترض هذه النظرية (ITR) إمكانية التنبؤ بأداء المفحوصين أثناء

الاستجابة على فقرات الاستبانة من خلال تحديد خصائص المفحوصين، أو ما يعرف بسماتهم أو قدراتهم على افتراض وجود متصل للسمة، بحيث يمكن تقدير احتمال الإجابة الصحيحة للفرد عن فقرة ما إذا علم موقعه (θ) على متصل السمة. وأن العلاقة بين أداء الفرد على الفقرة وقدرته يمكن أن تحدد من خلال ما يسمى بمنحنى خصائص الفقرة (Curve Characteristic Item)، وتفترض أيضاً أن مقدار الاحتمال يكون دالة متزايدة مطردة (Increasing Monotonically) لموقع الفرد على متصل السمة، مما يعني أن احتمال الإجابة الصحيحة يزداد بزيادة قدرة الفرد (Hambleton, 1994).

ولعلّ أحد أهم المزايا المرتبطة بنظرية الاستجابة للفقرة استقلالية القياس، ويعني أن تقدير معالم الفقرات يكون مستقلاً عن خصائص الأفراد التي استخدمت في تقدير هذه المعالم (Sample Free)، وأن تقدير قدرة الفرد يكون مستقلاً عن عينة الفقرات التي تطبق عليه (Hambleton, 1994) (Free-Item) وتعد استقلالية القياس بمثابة النقطة المفصلية بين النظرية الكلاسيكية ونظرية الاستجابة للفقرة، وهو ما عبر عنه لورد (Lord, 1980) بخاصية التغيير (الثبات) (Invariance). فإذا ما تحققت مثل هذه الخاصية يصبح بالإمكان القيام ببعض التطبيقات مثل: معادلة الاختبارات، وبنوك الأسئلة، والكشف عن تحيز الفقرات، والقياس التكميلي (Rup & Zumbo, 2006). بالإضافة إلى أن نظرية الاستجابة للفقرة تفترض وجود دالة لكل فقرة خاصة بها. ومعالم الفقرة هي: معالم الصعوبة (b_i): وهي تقدير القدرة المناظر لاحتمال الإجابة الصحيحة، ومعلمة تمييز الفقرة (a_i): تعبر عن الدرجة التي تستطيع بواسطتها الفقرة التمييز بين مستويات القدرة المختلفة. ومعالم تخمين الفقرة (C_i): هو الاستواء السفلي لمنحنى خصائص الفقرة (Hambleton & Swaminathan, 1985) وهذا المعلم خاص بالفقرات الموضوعية ثنائية التدرج من نوع الاختيار من متعدد، وهذا المعلم لا يخص بيانات الاستجابة متعددة التدرج.

وبنيت نظرية الاستجابة للفقرة (IRT) على مجموعة من الافتراضات ذكرها هامبلتون وسواميثان (Hambleton & Swaminathan, 1985) وهي: افتراض أحادية البعد (Unidimensionality) الذي يعني بأن يقيس الاختبار سمة واحدة تفسر أداء الفرد عليها، وافتراض الاستقلال الموضعي (Local Independence) ويقصد به أن تكون استجابات الفرد على فقرات الاختبار مستقلة إحصائياً عند مستوى قدرة معينة، أي أن استجابة الفرد عن فقرة ما يجب أن لا تؤثر سلباً أو إيجاباً على استجابته، وتقوم نماذج استجابة الفقرة على تحديد العلاقة بين أداء المفحوصين الملاحظ (Observable) على الأداة والسمات غير الملاحظة (Unobservable) المتوقعة للأداء على الاختبار، حيث وصف هذه العلاقة بدوال رياضية تقوم على افتراضات محددة فيما يخص بيانات الاختبار (Hambleton & Swaminathan 1985). وقد تم تطوير هذه النماذج كطرق لتحليل الاختبارات التربوية، وطُبقت بشكل واسع في المجالين النفسي والاجتماعي، إضافة إلى توافقها بشكل جيد مع نتائج مقاييس الاتجاهات (Fontanella, Villano & Donato, 2015).

وقد عُيّنت نظرية الاستجابة للفقرة بالقياس الموضوعي عنايةً دقيقة، فقد طورت النماذج الاحتمالية لتحديد العلاقة بين أداء الفرد على النماذج ثنائية التدرج ومنها النموذج

اللوغستي أحادي المعلم (One Parameter Logistic Model)، والنموذج اللوجستي ثنائي المعلم (Two Parameters Logistic Model)، والنموذج اللوجستي ثلاثي المعلم (Three Parameters Logistic Model) (Hambleton & Swaminathan, 1985).

وبالإضافة لذلك فقد عنيت النظرية بالنماذج الاحتمالية متعددة التدرج (Polytomous ITR Models ومنها نموذج الاستجابة المتدرجة المعدل (MGRM: Modified Graded Response Model)، ونموذج التقدير الجزئي العام (GPCM: Generalized Partial Credit Model)، ونموذج التقدير الجزئي (PCM: Partial Credit Model)، وأشهر النماذج متعددة التدرج؛ نموذج الاستجابة المتدرجة (Graded Response Model GRM)، وهو امتداد للنموذج ثنائي المعلم، ويفترض أن الاستجابات مرتبة وفقاً لمستوى موافقة المستجيب على نص الفقرة، ويناسب هذا النموذج المقاييس التي تصمم وفق مقياس ليكرت وهو النموذج الأكثر ملاءمة للتحقق من موضوعية أداة الاستبانة (Samejima, 1997)، ويتوافق ترتيب فئات الاستجابة في هذا النموذج مع بناء الفقرات في مقياس ليكرت، حيث فئات الاستجابة مرتبة من الأدنى إلى الأعلى تبعاً لقياسها لقوة الاتجاه نحو السمة (مثلاً: تعطي نقطة واحدة للاستجابة غير موافق بشدة، ونقطتان للاستجابة غير موافق، وثلاث نقاط للاستجابة موافق، وأربع نقاط للاستجابة موافق بشدة)، حيث إن الفرد الذي يختار الفئة الأعلى، يمتلك اتجاهًا إيجابيًا أكبر نحو السمة المقاسة أي (bi1 < bi2 < bi3 < bi4) (Samejima, 1997).

كما أن الاستجابة المتدرجة لساميجما من نماذج الفرق (Difference Models) التي يستخدم فيها الطرح للحصول على احتمالية الإجابة لفئة معينة، وهي مناسبة للاستجابات الرتببة، ويعد نموذج الاستجابة المتدرجة أول توسعة لنظرية استجابة الفقرة إلى الفقرات ذات فئات الإجابة المتعددة، واستخدمت ساميجما نماذج المنحنى الطبيعي واللوغستي كودات (Lustina, 2004)، إذ طورت ساميجما هذا النموذج (Samejima, 1994) ليتلاءم مع الإجابات على فقرة تحتوي على أكثر من فئتين رتبيتين للإجابة، وصنفت الاستجابات على الفقرة j إلى عدد $m_j + 1$ من الفئات الرتببة، وأعطيت درجات الفئة لفقرة j أعداداً صحيحة متتابعة، ويوجد في نموذج الاستجابة المتدرجة لساميجما (GRM) معلم تمييز ومجموعة من حدود الفئة لكل فقرة. وحددت ساميجما مرحلتين للحصول على احتمالية أن شخص ما بمستوى θ معين سيجيب على فئة معينة:

المرحلة الأولى: الاحتمالية لأي شخص ليجيب على فئة K أو أعلى فقرة j هي:

$$P_{K+}^*(\theta) = \frac{\exp[\alpha_i(\theta - (b_i - c_j))]}{1 + \exp[\alpha_i(\theta - (b_i - c_j))]}$$

حيث:

حيث: P_{ix}^* : احتمالية حصول المفحوص ذي القدرة (θ) على علامة فئة أكبر من أو تساوي (x) على الفقرة (i) ، α_i : معلم التمييز للفقرة (i) ، b_i : معلم الموقع (الصعوبة)، c_j : معلم عتبة الفئة.

المرحلة الثانية: الاحتمالية للإجابة على فئة معينة $P_{jx}(\theta)$ (احتمال الاستجابة في فئة x ضمن الفقرة i) بإجراء عملية طرح (θ) $p^*_{i(x+1)}$ (احتمال الاستجابة في الفئة $x+1$ أو أعلى) من $p^*_{ix}(\theta)$ (احتمال الاستجابة في الفئة x أو أعلى) كما في المعادلة:

$$P_{ix}(\theta) = P^*_{ix}(\theta) - P^*_{i(x+1)}(\theta)$$

ومن دقة اهتمام النظرية الحديثة للقياس بالحصول على القياس السيكومتري الموضوعي، فلم تقتصر النظرية على تطوير النماذج الإحصائية الاحتمالية، بل تعدت ذلك إلى إيجاد النماذج الإحصائية التي تعنى بالتحقق العددي لتقدير معالم الفقرات والأفراد للوصول إلى الهدف من بناء الأداة، وذلك عن طريق قياس المتغيرات التي لا يمكن قياسها مباشرة والتي تسمى بالسّمات أو القدرات الكامنة، بافتراض أن لكل مفحوص درجة قدرة تُعرف (Ability Scores)، والتي تستخدم في تفسير أداء كل من الفقرة والاختبار (De Ayala, 2009). فتقدير معالم الفقرة والأفراد عند تطبيق الاستبانة؛ يعد قضية مهمة عند استخدام نظرية الاستجابة للفقرة، خاصة وأن معظم البحوث التربوية التي تستخدم أداة الاستبانة تقوم على العينات العشوائية. وهنا يفترض المتخصصون بتطوير أدوات القياس أن السّمات تتوزع توزيعاً طبيعياً في المجتمع، وكلما كان حجم العينة كبيراً كانت العينة ممثلة، وبالتالي يكون توزيع السمة في العينة قريباً من التوزيع الطبيعي (Hambleton & Sawminthan, 1985).

ومن هنا فقد اهتمت تربويات القياس بالعوامل المؤثرة على دقة تقديرات القدرة، فمنهم من يرى لحجم العينة وطول الاختبار أثر على دقة التقدير؛ كما جاء في دراسة دراسغو (Drasgow, 1989) أن عملية التقدير تتأثر وبشكل كبير بحجم العينة وطول الاختبار، أما دراسة ستون (Stone, 1992) فقد أوضحت أن طول الاختبار له تأثير كبير ومباشر على تقدير معالم الصعوبة والقدرة سواء أكانت العينة صغيرة أم كبيرة. أما ستوكنج (Stocking, 1990) فبين أنه من المحتمل الحصول على تقديرات لمعالم الفقرة بشكل أدق إذا كان التخطيط لاختبار عينة المغايرة يأخذ بعين الاعتبار دقة تقدير معالم الفقرة.

ويشير هامبلتون وكوك (Hambleton & Cook, 1983) إلى أن عدد المفحوصين المستخدمين في تقدير معالم الفقرة وعدد الفقرات في الاختبار يؤثران على استقرار ودقة تقدير معالم الفقرات الثنائية، كما أنه من الممكن أن يكون لعدد فقرات الاختبار تأثير على ثبات وصدق الاختبار. وفي هذا السياق بين لورد (Lord, 1980) بشكل واضح أن طول الاختبار وحجم عينة المفحوصين يؤثران على جودة تقدير المعالم. وأشار ري وجينسن (Ree & Jenesen, 1983) إلى أن حجم العينة يؤثر على نتائج معادلة الاختبارات. وأثبتت دراسة كل من (بني عطا، Hambleton & Jones, 1980; Hambleton & Cook, 2013)

تقديرات معالم الفقرات وقدرة الأفراد عند استخدام نظرية الاستجابة للفقرات. (1994; Sass, Schmitt & Walker, 2004) بأن حجم العينة يلعب دوراً كبيراً في دقة

وأشار عيابة (2004) إلى أن الدقة في تقدير معالم القدرة تزداد عند زيادة عدد فقرات الاختبار، كما أن تقديرات معالم القدرة تكون أكثر دقة عند استخدام الطريقة المعيارية في اختيار الفقرات مقارنة بالطريقة العشوائية. ويشير هامبلتون وسوامينثان (Hambleton & Swaminathan, 1985) أن الاستبيان القصير من المتوقع أن تقدم تقديرات أكثر دقة عند مستويات القدرة المتدنية؛ لذا فهي أكثر فائدة للمفحوصين ذوي القدرة المتدنية، كما أن الاستبانة الطويلة من المتوقع أن تقدم معلومات أكثر دقة عند مستويات القدرة العليا، لذا فهي أكثر فائدة وفعالية للمفحوصين ذوي القدرة العليا؛ حيث يتضح بأن العينات المختلفة وطول الاختبار قد تولد تقديرات مختلفة.

ومن العوامل المؤثرة على دقة التقدير شكل توزيع القدرة؛ فبحسب ستون (1992) أن تقديرات معلمة الصعوبة بقيت مستقرة عبر توزيعات القدرة، ولكن تغيرت تقديرات معلمة التمييز عندما كان التوزيع المسبق للقدرة غير طبيعي، وبقيت مستقرة عند التوزيع الطبيعي للقدرة. وكشفت دراسة سونج (Seong, 1990) بأن تقديرات معلمة التمييز تكون أكثر دقة إذا كان شكل توزيع القدرة الحقيقية ملتوياً التواء سالباً وتوزيع شكل القدرة المسبق طبيعياً عندما يكون حجم العينة صغيراً، في حين كانت تقديرات معلمة الصعوبة أكثر دقة عندما كان شكل توزيع القدرة الحقيقية ملتوياً التواء موجباً وحجم العينة صغيراً. وأشار ساس وشميت ووكسر (Sass, Schmitt & Walker, 2004) بأن التقديرات في معالم الفقرات كانت أقل دقة عندما كان توزيع القدرة ملتوياً وأكثر دقة عندما كان توزيع القدرة طبيعياً، وعند مقارنة الأخطاء المعيارية بين معلمة الصعوبة والتمييز لوحظ بأنها الأعلى عند معلمة التمييز. كما يشير كل من بني عطا والشرفين (2011) أن تقديرات معالم الفقرات المعايير تبعا لشكل توزيع القدرة كانت دالة إحصائياً كمؤشر لاستقرار معالم الفقرات عبر التوزيعات المختلفة للقدرة. بالإضافة إلى أن دالة المعلومات للاختبار قد تباينت باختلاف شكل توزيع القدرة. ولكن يلاحظ أن جميع الدراسات السابقة الذكر مطبقة على الاختبارات التحصيلية ثنائية التدرج.

ومن الدراسات التي تناولت النماذج متعددة التدرج والتي عنيت بالبحث بالعوامل والظروف المؤثرة على دقة التقدير

فقد هدفت تدراسة جوليا (Golia, 2011) إلى تقييم أثر حجم الاستبانة على دقة الثبات باستخدام نموذج راش، وقد استخدمت الدراسة استبانة ذات مقياس متدرج خماسي مقابل كل فقرة، وأظهرت الدراسة ارتفاعاً في الثبات الإمبريقي إذ بلغ (0.95) في حال الاستبانة القصيرة، وأظهرت النتائج أن هناك علاقة عكسية بين طول الاستبانة ودقة تقدير واستقرار القدرة ولصالح طول الاستبانة القصيرة. كما وهدفت دراسة براون أوليفارز (Brown & Olivares, 2011) إلى الكشف عن فاعلية نظرية الاستجابة للفقرات في تقدير الاستجابة على الاستبانة ذات الاستجابات المحددة ومتعددة الأبعاد، وخفض المشكلات التي تواجه الطريقة التقليدية للقياس على فقرات

الاستبانة، من خلال دراسة محاكاة باستخدام بيانات محصلة من استبانة مكونة من (5) أبعاد صادرة عن المجموعة الدولية للفقرات الشخصية ل (438) مستجيباً، أظهرت الدراسة أن نظرية الاستجابة للفقرات قادرة على تقدير قدرات الأفراد والفقرات للاستبانة متعددة التدرج ومتعددة الأبعاد، كما أن استخدام نظرية الاستجابة للفقرات يقلل من احتمالية التحيز في الاستجابة، وأن دقة تقدير السمة تكون نتائجها أدق، وأظهرت النتائج ارتفاعاً في الخصائص السيكمترية للاستبانة على الفقرات وعلى الأبعاد. في حين أظهرت نتائج دراسة جانسن وآخرين (Jansen, et al. 2010) والتي هدفت إلى التحقق من الخصائص السيكمترية لاستبانتي لتقييم نوعية الحياة ذات الصلة بالمرضى المصابين بالتصلب المتعدد والشلل في الأطراف السفلية، أن معاملات ثبات كرونباخ ألفا كانت أعلى في نموذج الاستبانة (18) فقرات. مقارنة بنموذج (36) فقرات، وانخفض معامل ثبات كرونباخ ألفا في النموذج (136) فقرات. وأظهرت معاملات الصدق ارتفاعاً في نموذج (36 و 18) فقرات فيما انخفض لنموذج (136) فقرات. كما أظهرت معالم الفقرات من خلال النموذج اللوجستي فروقاً دالة إحصائياً لصالح النموذج ذو (18) فقرات. وهدفت دراسة إيلوت وهابدي (E.Bolt & Heide, 2013) إلى التحقق من تأثير طول الاستبانة عبر البريد الإلكتروني على معدلات الاستجابة في مسح طبي، ولتحقيق أهداف الدراسة تم إرسال استبانة اتخاذ قرار في نهاية العمر، مكونة من 4 صفحات إلى (1100) مشارك ومشاركة منهم (400) من أطباء رعاية المسنين، و(500) أخصائي طبي، كما تم إرسال استبانة إلكترونية أخرى أقصر مكونة من صفحتين إلى (500) أخصائي طبي، وبعد ثلاثة شهور تم إرسال استبانة مكونة من صفحة واحدة إلى الأطباء والأخصائيين الذين لم يستجيبوا على الاستبانتي السابقتين (4 صفحات والصفحتين). وأظهرت النتائج أن معدلات الاستجابة كانت متقاربة بين الاستبانتي (4 صفحات والصفحتين). حيث بلغ عدد المستجيبين (190) و(191) للنموذجين، فيما ارتفع مستوى معدل الاستجابة من (53%) إلى (64%) بعد إرسال النموذج القصير ذي الصفحة الواحدة من الاستجابة.

ومن الدراسات التي اهتمت بموضوع طول الاختبار دراسة بني عطا (2013)، التي هدفت إلى تقصي دقة تقدير النموذج اللوجستي ثلاثي المعالم لمعالم القدرة وقدره الأفراد، في ضوء تغير طول الاختبار وحجم العينة. ولتحقيق الهدف من الدراسة ولدت بيانات ثنائية الاستجابة بواقع 50 مرة لستة مستويات من طول الاختبار (10، 25، 50، 75، 100، 300 فقرات) وستة مستويات لحجم العينة (100، 250، 500، 1000، 2000، 4000) من خلال استخدام برنامج (WINGEN). وباستخدام برنامج (BILOG-MG3) حللت البيانات المولدة لكل خلية من تقاطع مستوى طول الاختبار ومستوى حجم العينة من أجل تقدير معالم الفقرات وقدره الأفراد، وإيجاد قيم مؤشر التحيز (Bias) والجذر التربيعي لمتوسط مربعات الخطأ (RMSE) لمعالم الفقرات وقدره الأفراد. وكشفت نتائج الدراسة عن وجود أثر ذي دلالة إحصائية لكل من طول الاختبار وحجم العينة والتفاعل بينهما في دقة تقديرات معالم الفقرات وقدره الأفراد. وكشفت النتائج أيضاً أن الوسط الحسابي لقيم RMSE لمعالم الفقرات وقدره الأفراد أخذ بالتناقص عندما زاد طول الاختبار على 50 فقرات وزاد حجم العينة على 2000. وقد أيدت قيم معاملات الارتباط شبه التام، بالإضافة إلى ذلك كانت قيم التحيز في التقديرات قريبة من الصفر.

وأجرت الحوارية (2014) دراسة هدفت للكشف عن أثر طول الاختبار وشكل توزيع القدرة في تقديرات القدرة للأفراد، وتقديرات معالم صعوبة الفقرات، واقتراح المعلومات للفقرات، وفق نموذج راش في نظرية استجابة الفقرة. ومن ضمن إجراءات الدراسة تم توليد البيانات، ومجموعة الفقرات، وقدرات الأفراد؛ إذ تم توليد بيانات بواقع (1000) مفحوص، بطول الاختبار (30، 60) فقرة وشكل توزيعات القدرة (ملتو التواء سالباً، طبيعياً، ملتو التواء موجباً)، وفق النموذج الأحادي المعالم. تم تحليل البيانات المولدة باستخدام البرمجيات (BILOG-MG3) و(WINGEN) و(SPSS). وأظهرت النتائج وجود فروق ذات دلالة إحصائية ($\alpha=0.05$) بين متوسطات الأخطاء المعيارية في تقديرات معالم قدرة الأفراد تعزى لشكل توزيع القدرة وطول الاختبار، حيث كانت تقديرات قدرة الأفراد أكثر دقة لصالح شكل التواء السالب والموجب. ولطول الاختبار (30) فقرة. كما أشارت النتائج إلى وجود فروق ذات دلالة إحصائية في تقديرات معالم الصعوبة، وكانت النتائج لصالح الاختبار بطول (30) فقرة؛ إذ زود بمعلومات أكثر. وتوصي الباحثة بإجراء المزيد من البحوث باعتماد النماذج المعلمية الأخرى، واستخدام بيانات حقيقية.

وأجرى ألين (Allen, 2016) دراسة حول تأثير اختصار طول الاستبانة على معدل الاستجابة وجودة الاستجابة؛ حيث اعتبر استبانة الخريجين لجامعة بيرمنغهام يونغ دراسة استقصائية أولية طويلة تحتوي على أكثر من 200 سؤال. وبعد إعداد نسخة مختصرة من الاستبانة، أظهرت زيادة معدلات الإجابة بصورة جوهرية، وكان المستجيبون من الذكور أكثر اقتناعاً بالإجابة على النسخة المختصرة مقارنة بالنسخة الطويلة.

وفي دراسة أجراها القيسي (2016) هدفت الكشف عن أثر حجم العينة وطول الاختبار على دقة تقدير معالم الفقرة والقدرة باستخدام نماذج نظرية الاستجابة للفقرة اللامعلمية، اعتماداً على مؤشري دقة القياس التحيز (BIAS)، والجذر التربيعي لمتوسط مربعات الخطأ (RMSE)، وقد أظهرت نتائج الدراسة وجود فروق دالة إحصائية دقة تقدير معلمة التمييز a تعزى لمتغير حجم العينة، كما أظهرت النتائج وجود فروق دالة إحصائية ($\alpha = 0.05$) في متوسطات مؤشر دقة القياس (BIAS) في دقة تقدير معلمة التخمين c تعزى لمتغيري الدراسة (حجم العينة، وطول الاختبار) ($\alpha = 0.05$) في متوسطات مؤشر دقة القياس الاختبار والتفاعل بينهما، كما بينت النتائج وجود فروق دالة إحصائية في مؤشر (BIAS) في دقة تقدير معلمة القدرة θ تعزى لمتغيري (حجم العينة، وطول الاختبار)، في حين لم تظهر فروق دالة إحصائية ($\alpha = 0.05$) في متوسطات مؤشر دقة القياس (BIAS) في دقة تقدير معلمة الصعوبة b .

بعد استعراض الدراسات السابقة يلاحظ أن بعض الدراسات هدفت إلى التعرف على الفروق بين الخصائص السيكومترية للنماذج متعددة التدرج (الاستبانة) في ضوء طول الاستبانة (عدد الفقرات) كدراسة كل من ولروماير وأوهناين (Waller, & Meyer, & Ohanian, 2001)، جوليا (Golia, 2011)، براون وأوليفارز (Brown & Olivares, 2011)، جانسن وآخرين (Jansen, et al. 2010) ومن الدراسات من استخدمت النظرية الحديثة للقياس في تقدير الخصائص السيكومترية للنماذج متعددة التدرج المتعلقة بالاستبانة والمقارنة مع النظرية التقليدية

في القياس كدراسة كل من جوليا (Golia, 2011)، براون واوليفارز (Brown & Olivares, 2011).

ومن خلال استعراض الدراسات السابقة يلاحظ أن الدراسات الأجنبية التي عُنيت بالتحقق من العوامل المؤثرة على النماذج متعددة التدرج، ويلاحظ ندرة الدراسات العربية والمحلية التي بحثت بشكل صريح بتقصي العوامل المؤثرة على دقة التقدير للنماذج متعددة التدرج، ومنها العلاقة بين طول الاستبانة على كل من الدقة في تقدير القدرة، وفق النظرية الحديثة للقياس، والخصائص السيكمترية للفقرات والمقياس، فأغلب الدراسات العربية كانت تتناول النماذج ثنائية التدرج (الاختبارات التحصيلية) في تقصي العوامل المؤثرة على دقة تقديرات القدرة. إن ما يميز هذه الدراسة أنها تعتبر بداية لتقصي العوامل المؤثرة على دقة تقديرات القدرة على النماذج متعددة التدرج في مجال القياس التربوي- في حدود علم الباحثين-.

مشكلة الدراسة وأسئلتها

تعد الاستبانة أكثر أدوات القياس شيوعاً في قياس السمات الشخصية والنفسية واستطلاعات الرأي، كما أنها أصبحت تُستخدم في شتى مجالات العلوم (الإدارية والصحية والاجتماعية والاقتصادية) وهذا يؤكد على أهمية الحصول على استبانة ذات موثوقية وموضوعية عالية للتمكن من الوثوق بالقرارات التي ستُتخذ بناء على نتائجها، بالإضافة إلى أن أداة الاستبانة في معظم البحوث تطبق على عينات ومجموعات كبيرة، إلا أن عملية جمع البيانات في البحوث التربوية من خلال الاستبانة قد يواجه بعض الصعوبات التي تجعل منها عملية تكاد تكون صعبة وغير مقبولة من أغلب المستجيبين، إذ قد يكون السبب طول فقرات الاستبانة المعدة في البحوث، شعور المستجيب بالملل أثناء الاستجابة خاصة بفقرات الاستبانة الطويلة، كما أن المستجيب إذا ما طبقت عليه الاستبانة في أوقات دوامه يشعر بالاستياء بأن فترة الراحة التي يمكن أن يحصل عليها أثناء فترة العمل ستذهب في الإجابة على الاستبانة، ويشعر باستياء أكبر إذا ما وجدها ذات فقرات طويلة، وهذا بدوره يشكل عاملاً هاماً في إمكانية عدم الحصول على دقة في الاستجابة من قبل المستجيب وستكون عملية التقدير عشوائية من قبله، وبالتالي الاحتمالية الكبرى الحصول على الأخطاء في القياس التي قد تؤثر على خصائصها السيكمترية، وعدم الحصول على دقة في البيانات وفي قياس تقدير معالم الفقرات والأفراد.

وفي الجانب الآخر نلاحظ بأنه في النظرية الكلاسيكية صيغة معادلة سبيرمان- براون تشير بأن زيادة عدد الفقرات يؤدي إلى زيادة الثبات، وكذلك كمية المعلومات في نظرية الاستجابة IRT تزداد بزيادة عدد الفقرات، ولكن يبقى السؤال ما هو عدد الفقرات الأنسب الذي يظهر من خلاله دقة التقديرات للقدرة ومعالم الفقرات.

ونظراً لأهمية تقدير القدرة للأفراد ومعالم الفقرات في القياس السيكمترية، فقد جاءت هذه الدراسة لتقصي العوامل المؤثرة في تقدير القدرة على النماذج متعددة التدرج وهي طول الاستبانة؛ ومن هنا جاءت الحاجة للقيام بعمل يثري عملية القياس السيكمترية للنماذج متعددة التدرج من

خلال دراسة أثر طول الاستبانة على دقة تقديرات القدرة والخصائص السيكمترية للفقرة والمقياس، وبالتحديد فإن هذه الدراسة تسعى للإجابة عن السؤالين التاليين:

- السؤال الأول: هل تختلف دقة تقدير معالم القدرة للأفراد باختلاف طول الاستبانة (10، 30، 60)؟
- السؤال الثاني: هل تختلف دقة تقدير معالم الصعوبة والتمييز لل فقرات باختلاف طول الاستبانة (10، 30، 60)؟

أهداف الدراسة

هدفت الدراسة الحالية التعرف إلى أثر طول الاستبانة على دقة تقدير معالم القدرة والخصائص السيكمترية للفقرة والمقياس في ضوء نظرية الاستجابة للفقرة، كما هدفت التعرف على طرق بناء استبانة مثالية كأداة قياس المتغيرات النفسية التي تمتاز بعدم الثبات النسبي، كونها تتأثر بالشمسية الإنسانية، كما تهدف الدراسة التعرف على أهم العوامل التي يجب أن تؤخذ بعين الاعتبار عند بناء الاستبانة للحصول على النتائج المرجوة من الدراسات النفسية والتربوية.

أهمية الدراسة

للدراسة أهمية نظرية وعملية على حد سواء، إذ تبرز الأهمية النظرية لهذه الدراسة في أنها تمثل إضافة للدراسات العربية التي ستبحث في موضوع طول الاستبانة، نظراً لكونها من الدراسات الأولى في هذا المجال – إن لم تكن الأولى من نوعها في حدود علم الباحثين- مما يساعد في إثراء الأدب التربوي، ويساعد في سد الفجوات في مجال الدراسات، وذلك من خلال الكشف عن الطول الأنسب لعدد فقرات الاستبانة (10، 30، 60)، في إعطاء تقديرات أدق للقدرة، أو الحصول على الخصائص السيكمترية الأفضل.

ومن الناحية العملية فإن الدراسة تحت الباحثين الآخرين في هذا المجال على العمل على إجراء مجموعة من الدراسات ذات العلاقة في موضوع طول الاستبانة، كما يتوقع أن يكون لنتائجها صدق كبير في الجهات المهتمة في هذا المجال من منظمات حكومية أو خاصة وغيرها، من خلال الاختيار الأفضل لطول الاستبانة.

حدود الدراسة ومحدداتها

تتحدد الدراسة بثلاث نماذج من الأطوال لاستبانة الرضى الوظيفي وهي (10، 30، 60)، وتأثيرها على دقة تقدير معالم القدرة والخصائص السيكمترية للفقرة والاستبانة، كما تتحدد الحدود البشرية بعينة عشوائية من المعلمين في دولة قطر من الجنسيات القطرية وغير القطرية خلال الفصل الدراسي الثاني من العام 2015/2016، كما تتحدد نتائج الدراسة بالخصائص السيكمترية لأداة الدراسة (استبانة الرضى الوظيفي).

التعريفات الاصطلاحية والإجرائية

تقدير المعالم للفقرات والأفراد: هي عملية التعبير الكمي عن المعالم باستخدام برنامج IRTPRO.

دقة التقدير: هو تعبير يشير لجودة التقدير التي يميزها الاحتمالية الكبيرة في أن التقدير قريب من القيمة الحقيقية، وفي هذه الدراسة تم استخدام معدل مربعات الأخطاء أو الأخطاء المعيارية في التقدير للحكم على مدى الدقة.

معالم الأفراد: وهي معالم قدرات الأفراد، وهي مقدار ما يمتلكه الفرد من السمة التي يتم قياسها تبعاً لنماذج نظرية الاستجابة للفقرة المتعددة التدرج لساميجما (GRM).

دالة المعلومات للفقرة $I_i(\Theta)$: يُعد من المفاهيم الأساسية في نظرية الاستجابة للفقرة سواء كانت الفقرة ثنائية التدرج، أو تتطلب استجابات متعددة التدرج، وتساعد هذه الدالة في بناء أدوات قياس تجعل دقة القياس أو المعلومات أكبر ما يمكن، وتبين مدى مساهمة الفقرة في دالة المعلومات للاستبانة بشكل مستقل عن الفقرات الأخرى، وفي هذه الدراسة سيتم تقديرها باستخدام دالة الأرجحية العظمى.

دالة المعلومات للاستبانة $I(\Theta)$: وهي دالة رياضية تعبر عن مجموع دوال المعلومات لجميع فقرات الاستبانة عند مستوى معين من القدرة.

الطريقة والإجراءات

منهج الدراسة

تم استخدام المنهج الوصفي للكشف عن أثر طول الاستبانة على دقة تقديرات القدرة للفقرة وفقاً لنموذج الاستجابة المُتدرّج (Graded Response Model (GRM)). وذلك لمناسبة هذا المنهج وطبيعة هذه الدراسة وأهدافها.

مجتمع الدراسة

تكون مجتمع الدراسة من جميع المعلمين البالغ عددهم (13112) في المدارس المستقلة في دولة قطر من الجنسيات القطرية وغير القطرية خلال الفصل الدراسي الثاني من العام 2015/2016، بحسب إحصائيات التعليم السنوية الصادرة عن وزارة التعليم والتعليم العالي في دولة قطر.

عينة الدراسة

تم اختيار عينة الدراسة بالطريقة العشوائية؛ وذلك بتوزيع نماذج الاستبانة على المعلمين ورقياً، إذ بلغ عدد أفراد العينة (1425) معلماً ومعلمة، منهم (1038) معلماً و(387) معلمة؛ موزعين إلى (492) معلماً ومعلمة لنموذج الاستبانة (10 فقرات) و(497) معلماً ومعلمة لنموذج الاستبانة (30 فقرة) و(436) معلماً ومعلمة لنموذج الاستبانة (60 فقرة × ورقي)، وذلك كما هو مبين في جدول (1).

جدول (1): توزيع أفراد عينة الدراسة تبعاً للمرحلة التعليمية والجنس وطول الاستبانة.

الكلّي	الجنس		المؤهل العلمي	المرحلة التعليمية	طول الاستبانة
	أنثى	ذكر			
151	42	109	تربوي	ابتدائي	10 فقرات
17	11	6	غير تربوي		
126	5	121	تربوي	إعدادي	
82	5	77	غير تربوي		
74	6	68	تربوي	ثانوي	
42	2	40	غير تربوي		
492	71	421	المجموع		
69	30	39	تربوي	ابتدائي	30 فقرة.
24	13	11	غير تربوي		
113	18	95	تربوي	إعدادي	
66	17	49	غير تربوي		
182	35	147	تربوي	ثانوي	
43	25	18	غير تربوي		
497	138	359	المجموع		
140	110	30	تربوي	ابتدائي	60 فقرة.
31	18	13	غير تربوي		
68	4	64	تربوي	إعدادي	
14		14	غير تربوي		
116	37	79	تربوي	ثانوي	
67	9	58	غير تربوي		
436	178	258	المجموع		
1425	387	1038	المجموع الكلّي		

أداة الدراسة

بهدف تحقيق أهداف الدراسة؛ فقد تم استخدام استبانة الرضى الوظيفي التي تطبق على المعلمين في وزارة التعليم والتعليم العالي في دولة قطر في السنوات السابقة والمطورة من قبل (عبد الجواد، متولي؛ 1993) ذات الصلة بموضوع الدراسة، التي تكوّنّت من (60) فقرة موزّعة إلى ستة مجالات؛ هي: التقدير واحترام الذات وله (12) فقرة ذات الأرقام (1، 5، 13، 17، 20، 26، 30، 33، 40، 48، 55، 59) ويتضمن أربع فقرات سالبة الاتجاه هي (20، 30، 33، 48)، والانتماء وله (10) فقرات ذات الأرقام

(2، 8، 10، 18، 21، 36، 38، 46، 50، 60) ويتضمن أربع فقرات سالبة الاتجاه هي (18، 21، 36، 60)، وطبيعة العمل في ذاته وله (12) فقرة ذوات الأرقام (3، 7، 12، 15، 23، 29، 34، 39، 41، 45، 54، 56) ويتضمن ست فقرات سالبة الاتجاه هي (3، 12، 15، 39، 45، 54)، والتفاعل الإداري والمهني وله (10) فقرات ذوات الأرقام (4، 9، 19، 28، 32، 37، 43، 47، 53، 57) ويتضمن ثلاث فقرات سالبة الاتجاه هي (9، 37، 57)، والتفاعل الاجتماعي وله (10) فقرات ذوات الأرقام (6، 11، 14، 22، 25، 35، 44، 49، 51، 58) ويتضمن ثلاث فقرات سالبة الاتجاه (6، 14، 51)، ثم الأجور والمكافآت وله (6) فقرات ذوات الأرقام (16، 24، 27، 31، 42، 52) ويتضمن ثلاث فقرات سالبة الاتجاه هي (24، 27، 42).

تم إنتاج نموذجين من الاستبانة المكونة من (60) فقرة؛ حيث تألف النموذج الأول من (10) فقرات ذوات الأرقام الزوجية ابتداءً من الرقم (6) وبزيادة مقدارها (6) إلى رقم الفقرة وصولاً إلى رقم الفقرة (60) وهي (6، 12، 18، 24، 30، 36، 42، 48، 54، 60)، كما تألف النموذج الثاني من (30) فقرة ذوات الأرقام الزوجية ابتداءً من الرقم (2) وبزيادة مقدارها (2) إلى رقم الفقرة وصولاً إلى رقم الفقرة (60) وهي (2، 4، 6، 8، 10، 12، 14، 16، 18، 20، 22، 24، 26، 28، 30، 32، 34، 36، 38، 40، 42، 44، 46، 48، 50، 52، 54، 56، 58، 60).

وتتطلب الإجابة عنها أن يُحدّد المعلم درجة موافقته على ما جاء في الفقرة ذات الاتجاه الإيجابي من خمس درجات حسب طريقة ليكرت، وذلك على النحو الآتي: الدرجة (1) وتعني أعارض تمامًا، الدرجة (2) وتعني أعارض، الدرجة (3) وتعني غير متأكد، الدرجة (4) وتعني أوافق، الدرجة (5) وتعني أوافق تمامًا. وأن يُحدّد المعلم درجة موافقته على ما جاء في الفقرة ذات الاتجاه السلبي من خمس درجات حسب طريقة ليكرت، وذلك على النحو الآتي: الدرجة (1) وتعني أوافق تمامًا، الدرجة (2) وتعني أوافق، الدرجة (3) وتعني غير متأكد، الدرجة (4) وتعني أعارض، الدرجة (5) وتعني أعارض تمامًا.

خصائص الاستبانة في صورتها الأصلية

قام الباحثان باستخدام عدة طرق للتحقق من صدق وثبات مقياس الرضا الوظيفي للمعلمين؛ وهي: الصدق المنطقي بعرض الاستبانة على مجموعة من المحكمين والمتخصصين في القياس، وقد حصلت على نسبة مرتفعة من الاتفاق على ملائمة الاستبانة، وتم التحقق من الصدق البنائي باستخدام صدق التجانس الداخلي واحتساب معامل ارتباط بيرسون بين الفقرة والدرجة الكلية للاستبانة، وأظهرت القيم أن جميع المعاملات دالة إحصائياً عند مستوى (0.01)، والصدق التمييزي من خلال المقارنة الطرفية أظهرت نتائج أن الاستبانة (تميز تمييزاً واضحاً بين المستويات العليا والمنخفضة من الرضا الوظيفي. أما فيما يتعلق بالتحقق من معاملات الثبات فقد تم استخدام ثبات طريقة إعادة الاختبار والذي بلغ (0.87)، وطريقة كرونباخ ألفا الذي بلغ (0.85). وجميعها معاملات صدق وثبات مرتفعة.

التحقق من افتراضات نموذج الاستجابة المتدرج (Graded Response Model (GRM))
التحقق من افتراض أحادية البعد لنماذج استبانة الرضى الوظيفي لدى المعلمين في دولة قطر
 للتحقق من افتراض أحادية البعد لفقرات نماذج استبانة الرضى الوظيفي تم إجراء التحليل
 العامل الاستكشافي باستخدام برنامج SPSS v25.0 بالاعتماد على بيانات التطبيق النهائي لعينة
 الدراسة المستهدفة، وجدول (2) يبين نتائج التحليل العامل الاستكشافي لفقرات نماذج الاستبانة
 وفقاً لطول الاستبانة:
جدول (2): نتائج التحليل العامل الاستكشافي لفقرات نماذج الاستبانة وفقاً لطول الاستبانة.

الجدور الكامنه الاولى			المكون	طول الاستبانة
التيانين المفسر التراكمي %	التيانين المفسر %	القيمة		
41.00	41.00	4.10	1	10
55.51	14.51	1.45	2	
68.95	13.44	1.34	3	
24.71	24.71	7.41	1	30
32.04	7.33	2.20	2	
37.51	5.47	1.64	3	
21.88	21.88	13.13	1	60
29.53	7.64	4.59	2	
33.95	4.42	2.65	3	

يلاحظ من جدول (2) أنَّ فقرات نماذج الاستبانة وفقاً لمتغير طول الاستبانة تُحقَّق أحادية
 البعد بثلاثة مؤشرات، كما تم حساب تشبعات (Factor loadings) فقرات نماذج الاستبانة وفقاً
 لطول الاستبانة والأخطاء المعيارية لها (Cai, 2010-a, 2010-b, 2010-c) باستخدام برنامج
 IRTPro v4، وجدول (3) يُبين النتائج من إحصاءات القيمتين الصغرى والعظمى والوسط
 الحسابي والانحرافات المعيارية لتشبعات فقرات نماذج الاستبانة والأخطاء المعيارية لها وفقاً
 لطول الاستبانة.

جدول (3): إحصاءات القيمتين الصغرى والعظمى والوسط الحسابي والانحراف المعياري
لتشبعات فقرات نماذج الاستبانة والأخطاء المعيارية لها وفقاً لطول الاستبانة.

طول الاستبانة	الإحصائي	λ_1	الخطأ المعياري
10	القيمة الصغرى	0.534	0.053
	الوسط الحسابي	0.639	0.064
	الانحراف المعياري	0.072	0.008
	القيمة العظمى	0.736	0.074

...تابع جدول رقم (3)

طول الاستبانة	الإحصائي	λ_1	الخطأ المعياري
30	القيمة الصغرى	0.401	0.059
	الوسط الحسابي	0.691	0.076
	الانحراف المعياري	0.153	0.011
	القيمة العظمى	0.899	0.091
60	القيمة الصغرى	0.405	0.042
	الوسط الحسابي	0.595	0.071
	الانحراف المعياري	0.130	0.011
	القيمة العظمى	0.899	0.089

يلاحظ من جدول (3) وجود فروق ملاحظة بين الأوساط الحسابية لقيم تشبعات الفقرات وفقاً لطول الاستبانة.

التحقق من افتراض الاستقلال الموضعي لفقرات نماذج استبانة الرضى الوظيفي لدى المعلمين في دولة قطر

للتحقق من افتراض الاستقلال الموضعي لفقرات نماذج استبانة الرضى الوظيفي فقد تم حساب قيمة χ^2 للاستقلال الموضعي المعيارية (Standardized LD χ^2) لكل زوج متشكّل من تقاطع أي فقرتين من فقرات نماذج الاستبانة؛ باستخدام برنامج IRT Pro v4، فإذا كانت قيمة χ^2 للاستقلال الموضعي المعيارية أقل من (5) فمعنى ذلك أن تقاطع زوج الفقرتين يُحقّق الاستقلال الموضعي وإذا كانت قيمته أكبر من أو تساوي (5) فمعنى ذلك أن تقاطع زوج الفقرتين لا يُحقّق الاستقلال الموضعي، ثم تم رصد التكرارات والنسب المئوية لحالة الاستقلال الموضعي لكل تقاطع زوج فقرتين، وذلك كما هو مبين في جدول 4.

جدول (4): التكرارات والنسب المئوية لحالة الاستقلال الموضعي لفقرات نماذج الاستبانة وفقاً لطول الاستبانة.

طول الاستبانة	حالة الاستقلال الموضعي	تكرار تقاطع أزواج الفقرات	النسبة المئوية
10	تبعية موضعية	20	44.44
	استقلال موضعي	25	55.56
	الكلي	45	100.00
30	تبعية موضعية	153	35.17
	استقلال موضعي	282	64.83
	الكلي	435	100.00

...تابع جدول رقم (4)

طول الاستبانة	حالة الاستقلال الموضعي	تكرار تقاطع أزواج الفقرات	النسبة المئوية
60	تبعية موضوعية	121	6.84
	استقلال موضعي	1649	93.16
	الكلي	1770	100.00

يلاحظ من جدول (4) أنَّ الاستقلال الموضعي مُتحقق لنماذج الاستبانة وفقاً لطول الاستبانة حيث تراوحت النسب المئوية للاستقلال الموضعي من (55.56%) وحتى (100.00%).

التحقق من مطابقة معالم القدرة للمعلمين على فقرات نماذج استبانة الرضى الوظيفي لدى المعلمين في دولة قطر

تم استخدام برنامج IRTPro v4؛ لتقدير معالم القدرة للمعلمين على فقرات نماذج الاستبانة والخطأ المعياري لها واحتمالية الخطأ لمطابقة المعلمين، ثم تم حساب إحصاءات القيمتين الصغرى والعظمى والوسط الحسابي والانحراف المعياري لقدرات المعلمين المُقدَّرة في معرض استجاباتهم على فقرات نماذج الاستبانة والأخطاء المعيارية لها واحتمالية الخطأ لمطابقتهم وفقاً لطول الاستبانة، وذلك كما هو مبين في جدول 5.

جدول (5): إحصاءات القيمتين الصغرى والعظمى والوسط الحسابي والانحراف المعياري لقدرات المعلمين المُقدَّرة في معرض استجاباتهم على فقرات نماذج الاستبانة والأخطاء المعيارية لها واحتمالية الخطأ لمطابقتهم وفقاً لطول الاستبانة.

طول الاستبانة	الإحصائي	القدرة وفق EAP	الخطأ المعياري للقدرة	احتمالية الخطأ
10	القيمة الصغرى	-2.723	0.333	0.01088
	الوسط الحسابي	-0.000199	0.384817	0.38279
	الانحراف المعياري	0.9234	0.0305	0.29404
	القيمة العظمى	3.875	0.519	1.00000
30	القيمة الصغرى	-3.464	0.299	0.01013
	الوسط الحسابي	-0.000103	0.341491	0.38520
	الانحراف المعياري	0.9400	0.0368	0.29476
	القيمة العظمى	3.799	0.603	0.99783
60	القيمة الصغرى	-1.756	0.165	0.01043
	الوسط الحسابي	0.000989	0.207511	0.26940
	الانحراف المعياري	0.9786	0.0339	0.28384
	القيمة العظمى	3.42	0.429	0.97208

يلاحظ من جدول (5) أنَّ أكبر وسط حسابي لقدرات المعلمين المُقدَّرة على فقرات نماذج الاستبانة يكون عند طول الاستبانة (60) قد بلغت قيمته (0.000989) بانحراف معياري مقداره (0.9786) حيث تراوحت قيم قدرة المعلمين فيها من (-756.1) وحتى (3.42)، وتراوحت أخطاءها المعيارية من (0.165) وحتى (0.429)، وتراوحت احتمالية الخطأ لمطابقتهم فيها من (0.01043) وحتى (0.97208). مما يُعني أنَّ نموذج الاستبانة المتدرج GRM يُطابق جميع المعلمين وفقاً لطول الاستبانة.

التحقق من دلالات صدق نماذج استبانة الرضى الوظيفي لدى المعلمين في دولة قطر

للتحقق من دلالات صدق فقرات نماذج استبانة الرضى الوظيفي لدى المعلمين في دولة قطر تم حساب قيم معاملات الارتباط المُصحَّح لفقرات نماذج الاستبانة مع الاستبانة (Corrected Item-Total correlation) باستخدام برنامج SPSS v25.0، للتحقق من دلالات صدق فقرات نماذج الاستبانة وفقاً لطول الاستبانة. و جدول (6) يُبين إحصاءات القيمتين الصغرى والعظمى والوسط الحسابي والانحراف المعياري لقيم معاملات الارتباط المُصحَّح لفقرات مع نماذج الاستبانة التي تتبع لها وفقاً لطول الاستبانة:

جدول (6): إحصاءات القيمتين الصغرى والعظمى والوسط الحسابي والانحراف المعياري لقيم معاملات الارتباط المُصحَّح لفقرات مع نماذج الاستبانة التي تتبع لها وفقاً لطول الاستبانة.

طول الاستبانة	الإحصائي	قيم معاملات الارتباط المصححة
10	القيمة الصغرى	0.490
	الوسط الحسابي	0.531
	الانحراف المعياري	0.034
	القيمة العظمى	0.587
30	القيمة الصغرى	0.310
	الوسط الحسابي	0.564
	الانحراف المعياري	0.224
	القيمة العظمى	0.997
60	القيمة الصغرى	0.329
	الوسط الحسابي	0.491
	الانحراف المعياري	0.109
	القيمة العظمى	0.971

يلاحظ من جدول (6) أنَّ أصغر وسط حسابي لقيم معاملات الارتباط المُصحَّح لفقرات مع نماذج الاستبانة التي تتبع لها يكون عند طول الاستبانة (60) قد بلغت قيمته (0.491) وأنَّ أكبر

وسط حسابي لقيم معاملات الارتباط المُصحَّح للفقرات مع نماذج الاستبانة التي تتبع لها يكون عند طول الاستبانة (30) قد بلغت قيمته (0.564).

التحقُّق من ثبات نماذج استبانة الرضى الوظيفي لدى المعلمين في دولة قطر

تم حساب قيم الاتساق الداخلي باستخدام Cronbach's α ؛ للتحقق من الثبات الكلاسيكي لفقرات نماذج الاستبانة وفقاً لطول الاستبانة باستخدام برنامج SPSS v25.0، وبلغت قيم معامل ثبات كرونباخ ألفا (0.837) لنموذج الاستبانة (10) فقرات، و(0.748) لنموذج الاستبانة (30) فقرة، و(0.930) لنموذج الاستبانة (60) فقرة.

متغيرات الدراسة

اشتملت الدراسة على المتغيرات الآتية:

- أ. المتغيرات المستقلة؛ وهي:
 - طول الاستبانة: وله ثلاثة مستويات؛ هي: (10، 30، 60) فقرة.
- ب. المتغيرات التابعة؛ وهي:
 - تقديرات قدرات الأفراد.

المعالجات الإحصائية

للإجابة عن سؤال الدراسة الأول فقد تم حساب الأوساط الحسابية والأخطاء المعيارية وفترة الثقة لها والانحرافات المعيارية للخطأ المعياري في تقدير قدرات الأفراد وفقاً لطول الاستبانة، ثم تم إجراء تحليل التباين ثنائي التفاعل لها وفقاً لطول الاستبانة، ومن ثم إجراء اختبار ليفين (Levene) للكشف عن انتهاك تجانس التباين من عدمه بين الأوساط الحسابية للخطأ المعياري في تقدير قدرات الأفراد وفقاً لطول الاستبانة، ثم تم استخدام اختبار جيمس-هوبل (Games-Howell) للمقارنات البعدية المتعددة.

نتائج الدراسة ومناقشتها

النتائج المتعلقة بسؤال الدراسة الأول الذي نصَّ على: "هل تختلف دقة تقدير معالم القدرة للأفراد باختلاف طول الاستبانة (10، 30، 60)؟"

للإجابة عن سؤال الدراسة؛ فقد تم حساب الأوساط الحسابية والأخطاء المعيارية وفترة الثقة لها والانحرافات المعيارية للخطأ المعياري في تقدير قدرات الأفراد وفقاً لطول الاستبانة (10، 30، 60) وذلك كما هو مبين في جدول (7).

جدول (7): الأوساط الحسابية والأخطاء المعيارية وفترة الثقة لها والانحرافات المعيارية للخطأ المعياري في تقدير قدرات الأفراد وفقاً لطول الاستبانة.

طول الاستبانة	العدد	الوسط الحسابي	الخطأ المعياري	الانحراف المعياري	فترة الثقة 95%	
					الحد الأدنى	الحد الأعلى
10	492	0.385	0.002	0.030	0.382	0.388
30	497	0.341	0.002	0.037	0.338	0.345
60	436	0.208	0.002	0.034	0.204	0.211
الكلية	1425	0.315	0.001	0.081	0.309	0.313

يلاحظ من جدول (7) وجود فروق ظاهرة بين الأوساط الحسابية للخطأ المعياري في تقدير قدرات الأفراد ناتجة عن اختلاف مستويات طول الاستبانة. وللتحقق من جوهرية الفروق الظاهرة؛ فقد تم إجراء تحليل التباين الأحادي، وذلك كما هو مبين في جدول (8).

جدول (8): نتائج تحليل التباين الأحادي التفاعل للخطأ المعياري في تقدير قدرات الأفراد وفقاً لطول الاستبانة.

مصدر التباين	مجموع المربعات	درجة الحرية	وسط مجموع المربعات	ف	الدلالة الإحصائية	حجم الأثر
طول الاستبانة	16.274	2	8.137	6924.991	0.0001	82.32%
الخطأ	3.495	2974	0.001			
الكلية	20.659	2979				

يتضح من جدول (8) وجود فروق دالة إحصائية عند مستوى الدلالة ($\alpha=0.05$) بين الأوساط الحسابية للخطأ المعياري في تقدير قدرات الأفراد يُعزى لطول الاستبانة؛ ولكون طول الاستبانة متعدد المستويات، فقد تم إجراء اختبار ليفين (Levene) للكشف عن انتهاك تجانس التباين من عدمه بين الأوساط الحسابية للخطأ المعياري في تقدير قدرات الأفراد وفقاً لطول الاستبانة، حيث بلغت قيمة F المحسوبة لاختبار ليفين ما مقداره (7.89) بدلالة إحصائية عند مستوى الدلالة ($\alpha=0.05$) عند درجتي حرية (5) للبسط، و (2974) للمقام؛ بما يفيد وجود انتهاك في تجانس التباين؛ مما أوجب استخدام أحد اختبارات المقارنات البعدية التي تراعي انتهاك تجانس التباين حيث تم استخدام اختبار جيمس-هويل (Games-Howell) للمقارنات البعدية المتعددة؛ بهدف تحديد لصالح أي من مستويات طول الاستبانة قد كانت الفروق الجوهرية بين الأوساط الحسابية للأخطاء المعيارية في تقدير قدرات الأفراد، وذلك كما هو مبين في جدول 10.

جدول (9): نتائج اختبار جيمس-هويلل للمقارنات البعدية بين الأوساط الحسابية للأخطاء المعيارية في تقدير قدرات الأفراد وفقاً لطول الاستبانة.

طول الاستبانة				
10	30	60	الوسط الحسابي	Games-Howell
0.38	0.32	0.20		
0.18	0.12		0.20	60
0.06			0.32	30
			0.38	10

يتضح من جدول (9) أنَّ الفروق الجوهرية بين الأوساط الحسابية للخطأ المعياري في تقدير قدرات الأفراد وفقاً لطول الاستبانة؛ قد كانت لصالح طول الاستبانة (60) مقارنةً بكلِّ من طول الاستبانة (10) و(30) فقرة ثم لصالح طول الاستبانة (30) مقارنةً بطول الاستبانة (10).

أي أن طول الاستبانة المكون من 60 يعطي تقدير لقدرات الأفراد أفضل من طول الاستبانة 30 و10، وأن دقة التقدير كانت بشكل تنازلي تبعاً لطول الاستبانة، وقد تُعزى النتيجة إلى أن نظرية الاستجابة للفقرة قادرة على تحديد السمات الكامنة للمفحوصين، كما أن استخدام نظرية الاستجابة للفقرة في التحقق من دقة تقديرات الفقرة يقلل من احتمالية التحيز في الاستجابة، كما أن دقة تقدير السمة تكون نتائجها أدق. وهذا ما اتفقت عليه مع نتائج دراسة براون أوليفارز (Brown & Olivares, 2011)، وربما تُعزى النتيجة إلى أن من العوامل المؤثرة على دقة تقديرات الفقرة عدد الفقرات في الاستبانة، كما يشير هامبلتون وكوك (Hambleton & Cook, 1983) إلى أن عدد الفقرات في الاختبار تؤثران على استقرار ودقة تقدير معلمات الفقرات، كما أنه من الممكن أن يكون لعدد فقرات الاختبار تأثير على ثبات وصدق الاختبار. ويشير لورد (Lord, 1980) أن طول الاختبار يؤثر على جودة تقدير المعلمات. وقد تُعزى النتيجة إلى أن عدد فقرات الاختبار يؤثر على دقة تقدير قدرات الأفراد، فعلى الرغم من أن قدرات الأفراد تكون مستقلة عن فقرات الاختبار إلا أن أحد أهم العوامل المؤثرة عليها عدد فقرات الاختبار.

كما قد تُعزى النتيجة إلى أن طبيعة البيانات والمعلومات وعدد الفقرات تؤثر على كمية ودقة المعلومات التي تقدمها الاستبانة، خاصة وأن الاستبيان يقيس الرضى الوظيفي للمعلمين والذي يعتبر سمة هامة للمعلمين، ويريدون أن يعطوا أكبر قدر من المعلومات حول الرضى الوظيفي مما أثر على دقة التقدير على فقرات الاستبانة ذات 60 فقرة، لكون هذا العدد من الفقرات يحوي على كمية كبيرة من المعلومات التي تهتم بالمفحوصين. وهنا يفضل أن يؤخذ بعين الاعتبار موضوع وهدف بناء الاستبانة في تحديد عدد الفقرات التي يمكن أن تقدم أكبر قدر من المعلومات، وبالتالي الحصول على دقة تقدير عالية. بالإضافة لذلك ربما تُعزى النتيجة إلى أن طول الاستبانة له أثر على توازن أو تجانس المحتوى، وأن كمية المعلومات التي تعطيها تلك الفقرات (60) فقرة، كافية لتغطية متصل السمة، وإن زيادة حجم الفقرات يبقي المعلومات ثابتة وبالتالي يعطي دقة في تقدير الأفراد. واتفقت النتيجة مع نتيجة دراسة بني عطا (2014) أنه لإنتاج دقة تقديرات دقيقة يفضل استخدام أطوال تزيد عن 50 فقرة.

واختلفت النتيجة مع نتائج دراسة جوليا (Golia, 2011) التي أظهرت أن هناك علاقة عكسية بين طول الاستبانة ودقة تقدير واستقرار القدرة. ودراسة جانسن وآخرون (Jansen, et al. 2010) التي أظهرت وجود فروق دالة إحصائية للنماذج قصيرة الفقرات، واختلفت مع نتائج دراسة بيلوت وهايدي (E.Bolt & Heide, 2013) التي أظهرت ارتفاع مستوى معدل الاستجابة لنموذج الاستبانة القصير.

النتائج المتعلقة بسؤال الدراسة الذي نصّ على: "هل تختلف دقة تقدير معالم الصعوبة والتمييز للفقرات باختلاف طول الاستبانة (10، 30، 60)؟"

للإجابة عن سؤال الدراسة الثاني؛ فقد تم حساب الأوساط الحسابية والانحرافات المعيارية للأخطاء المعيارية في تقدير معالم التمييز للفقرات وفقاً لطول الاستبانة، وذلك كما هو مبين في جدول (10).

جدول (10): الأوساط الحسابية والانحرافات المعيارية للأخطاء المعيارية في تقدير معالم التمييز للفقرات وفقاً لطول الاستبانة.

طول الاستبانة	الوسط الحسابي	الانحراف المعياري
10	0.14	0.02
30	0.11	0.02
60	0.13	0.03
الكلي	0.12	0.03

يلاحظ من جدول (10) وجود فروق ظاهرة بين الأوساط الحسابية للأخطاء المعيارية في تقدير معالم التمييز للفقرات ناتجة عن اختلاف مستويات طول الاستبانة، وللتحقق من جوهرية الفروق الظاهرة؛ فقد تم إجراء تحليل التباين الأحادي التفاعل للأوساط الحسابية والأخطاء المعيارية في تقدير معالم التمييز للفقرات وفقاً لطول الاستبانة، وذلك كما هو مبين في جدول (11).

جدول (11): نتائج تحليل التباين الأحادي للأوساط الحسابية للأخطاء المعيارية في تقدير معالم التمييز للفقرات وفقاً لطول الاستبانة.

مصدر التباين	مجموع المربعات	درجة الحرية	وسط مجموع المربعات	ف	الدالة الإحصائية	حجم الأثر
طول الاستبانة	0.01	2	0.01	9.92	0.001	9.28%
الخطأ	0.14	194	0.00			
الكلي	0.15	199				

يتضح من جدول (11) وجود فروق دالة إحصائية عند مستوى الدلالة ($\alpha=0.05$) بين الأوساط الحسابية للخطأ المعياري في تقدير معالم التمييز للفقرات يُعزى لطول الاستبانة؛ ولكن

طول الاستبانة متعدد المستويات، فقد تم إجراء اختبار ليفين (Levene) للكشف عن انتهاك تجانس التباين من عدمه بين الأوساط الحسابية للأخطاء المعيارية في تقدير معالم التمييز للفقرات وفقاً لطول الاستبانة، حيث بلغت قيمة F المحسوبة لاختبار ليفين ما مقداره (1.26) بدون دلالة إحصائية عند مستوى الدلالة ($\alpha=0.05$) عند درجتي حرية (5) للبسط، و(194) للمقام؛ بما يفيد عدم وجود انتهاك في تجانس التباين؛ مما أوجب استخدام أحد اختبارات المقارنات البعدية التي تراعي تجانس التباين؛ حيث تم استخدام اختبار شفييه (Scheffe) للمقارنات البعدية المتعددة؛ بهدف تحديد لصالح أيٍّ من مستويات طول الاستبانة وقد كانت الفروق الجوهرية بين الأوساط الحسابية للأخطاء المعيارية في تقدير معالم التمييز للفقرات، وذلك كما هو مبين في جدول (12).

جدول (12): نتائج اختبار شفييه للمقارنات البعدية بين الأوساط الحسابية للأخطاء المعيارية في تقدير معالم التمييز للفقرات وفقاً لطول الاستبانة.

طول الاستبانة		10	30	60
الوسط الحسابي		0.14	0.11	0.12
10	0.14		0.01	0.03
30	0.11			0.02
60	0.12			

يتضح من جدول (12) أنَّ الفروق الجوهرية بين الأوساط الحسابية للأخطاء المعيارية في تقدير معالم التمييز للفقرات وفقاً لطول الاستبانة؛ قد كانت لصالح طول الاستبانة (30) مقارنةً بطول الاستبانة (10)، ثم لصالح طول الاستبانة (60) مقارنةً مع طول الاستبانة (10) في حين لم يكن هناك من فرقٍ دالٍ إحصائياً عند مستوى الدلالة ($\alpha=0.05$) بين الوسطين الحسابيين للأخطاء المعيارية في تقدير معالم التمييز للفقرات عندما يكون طول الاستبانة (60) مقارنةً مع طول الاستبانة (30).

وتعزى النتيجة إلى أن دقة تقدير القدرة التي تم الحصول عليها في طوال الاستبانة 60 و30 انعكس على قيم معاملات ارتباط معالم التمييز، إذ إن معالم التمييز هو بالأصل معاملات ارتباط. فطول الاستبانة من أهم العوامل المؤثرة على معامل ارتباط الفقرات. واختلفت مع نتائج دراسة جانسن وآخرين (jansen, et al, 2010) التي أظهرت فروق دالة إحصائية لمعامل الفقرات لصالح النموذج ذي (18) فقرة مقارنة مع نموذج (36) فقرة.

وكذلك للإجابة عن سؤال الدراسة الثاني؛ فقد تم حساب الأوساط الحسابية والانحرافات المعيارية للأخطاء المعيارية في تقدير معالم العتبة للفقرات وفقاً لمعالم العتبة (الأولى، الثانية، الثالثة، الرابعة) ولطول الاستبانة وذلك كما هو مبين في جدول (13).

جدول (13): الأوساط الحسابية والانحرافات المعيارية للأخطاء المعيارية في تقدير معالم العتبة للفقرات وفقاً لعتبة الصعوبة وطول الاستبانة.

طول الاستبانة	الخطأ المعياري لمعالم العتبة							
	الأولى		الثانية		الثالثة		الرابعة	
	الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري	الوسط الحسابي	الانحراف المعياري
10	0.16	0.04	0.09	0.02	0.14	0.03	0.31	0.08
30	24.03	97.44	15.30	64.13	8.14	34.43	5.54	18.02
60	0.96	2.35	0.45	1.03	0.37	1.07	0.94	2.72
الكلية	7.80	53.84	4.87	35.39	2.68	19.00	2.26	10.21

يلاحظ من جدول (13) وجود فروق ظاهرة بين الأوساط الحسابية للأخطاء المعيارية في تقدير معالم العتبة للفقرات ناتجة عن اختلاف عتبات الصعوبة وطول الاستبانة، وللتحقق من جوهرية الفروق الظاهرة؛ فقد تم إجراء تحليل التباين للقياسات المتكررة ثنائي التفاعل للأوساط الحسابية للأخطاء المعيارية في تقدير معالم العتبة للفقرات وفقاً لعتبة الصعوبة ولطول الاستبانة، وذلك كما هو مبين في جدول (14).

جدول (14): نتائج تحليل التباين للقياسات المتكررة ثنائي التفاعل للأوساط الحسابية للأخطاء المعيارية في تقدير معالم العتبة للفقرات وفقاً لمعالم العتبة ولطول الاستبانة.

أثر اختبارات	مصدر التباين	مجموع المربعات	درجة الحرية	وسط مجموع المربعات	ف	الدالة الإحصائية	حجم الأثر
داخل الأفراد [Mauchly's W=0.00، χ^2 التقريبية=2007.25، درجة الحرية = 5، الدالة الإحصائية = 0.00، ϵ (Greenhouse – Geisser) = 0.337]							
معالم العتبة	854.44	1.01	844.75	1.49	0.22	0.76%	
معالم العتبة × طول الاستبانة	1999.95	2.02	988.62	1.74	0.18	1.76%	
الخطأ (عتبة الصعوبة)	111585.9	196.23	568.66				
الكلية	117270.8	202.30	579.699				
بين الأفراد							
طول الاستبانة	6339.87	2	3169.93	1.78	0.17	1.80%	
الخطأ	345857.8	194	1782.77				
الكلية	361307.7	199	1815.62				
الكلية	478578.5	401.30	1192.58				

يتضح من جدول (14) عدم وجود فروق دالة إحصائية عند مستوى الدلالة ($\alpha=0.05$) بين الأوساط الحسابية للخطأ المعياري في تقدير معالم العتبة للفقرات يُعزى لمعالم العتبة.

فيما يتضح من جدول (14) عدم وجود فروق دالة إحصائية عند مستوى الدلالة ($\alpha=0.05$) بين الأوساط الحسابية للوسط الحسابي للخطأ المعياري في تقدير معالم العتبة للفقرات يُعزى لطول الاستبانة.

ويتضح من جدول (14) عدم وجود فروق دالة إحصائية عند مستوى الدلالة ($\alpha=0.05$) بين الأوساط الحسابية للخطأ المعياري في تقدير معالم العتبة للفقرات يُعزى لتفاعل معالم العتبة مع طول الاستبانة.

وتعزى النتيجة إلى أن السبب في عدم وجود فروق لعتبة الصعوبة ربما يعود لاختلاف تباين قدرة الأفراد على الرغم من أن الفقرات نفسها كانت لجميع الأفراد، وهذا يتفق مع ما أشارت إليه نتائج تحليل التباين الأحادي للقياسات المتكررة باختلاف معالم الصعوبة وطول الاستبانة بعدم وجود فروق تعزى لعتبة الصعوبة وطول الاستبانة والتفاعل بينهما. كما قد تعزى النتيجة إلى أن الحصول على دقة تقدير للأفراد يشير إلى أن الالتواء لشكل الفقرة إلى اليمين أو اليسار مما يدل على تقارب صعوبة الفقرة والتميز مما يقلل من قيمة الأخطاء المعيارية وبالتالي عدم وجود فروق إحصائية.

كما أن الأوساط الحسابية للأخطاء المعيارية جاءت متقاربة عند أطوال الاستبانة 60 لجميع تقديرات القدرة مما يدل على عدم وجود تباين في قدرات الأفراد في الإجابة عن هذه الفقرات. واختلفت النتيجة مع نتائج دراسة جانسن وآخرين (Jansen, et al. 2010) التي أظهرت فروق دالة إحصائية لمعالم الفقرات لصالح النموذج ذي (18) فقرة مقارنة مع طول (36) فقرة.

التوصيات

أشارت نتائج الدراسة إلى أن دقة تقديرات القدرة تتأثر بطول الاستبانة في ضوء نظرية الاستجابة للفقرة، وبناء على ذلك تم وضع التوصيات الآتية:

- على مطوري الاستبانات كأداة لبحوثهم مراعاة طول الاستبانة للحصول على دقة في تقديرات القدرة للأفراد عينة الدراسة. والأفضل أن تكون من ذات الأكثر من (30) فقرة.
- لفت انتباه المختصين بالقياس والتقويم لإجراء البحوث والدراسات التي تهدف إلى التحقق من الطرق والعوامل التي تؤثر على الخصائص السيكومترية للاستبانة ودقة تقديرات القدرة، وإعطائها العناية بالبحوث كنظيراتها من أدوات القياس وخاصة الاختبارات التحصيلية
- استخدام الاستبانات ذات الأطوال التي تتراوح بين (30-60) في البحوث التربوية.
- إجراء الدراسات التي تتناول أثر متغيرات أخرى على دقة تقديرات القدرة كنوع التدرج للاستبانة (خماسي، رباعي، ثلاثي)، وطبيعة السمة التي تقيسها الاستبانة،
- إجراء دراسات تبحث في أثر توزيع شكل القدرة على معالم الفقرات.

References (Arabic & English)

- Ababneh, I. (2004). *The Effect of sample size, method of sampling, number of items and method of item selection on the accuracy of estimation item and ability parameters for mental test utilizing Items*, Unpublished PhD thesis, Amman Arab University for Graduate Studies, Amman.
- Abdel-Gawad, N. & Metwally, M. (1993). The level of job satisfaction of teachers, *Journal of Educational Studies*, Association of Modern Education, 8(51), World of the Book, Cairo.
- Adas, A. (2003). *Psychometric measurement theory and application*.4, Cairo, dar alfeke eraharabi.
- Agresti, A. & Finnlay, B. (2009). *Statistical methods for the social science* (4th Ed). Upper Saddle River, NJ: Person- Prentice Hall.
- Al -Tal, W. & Qahel, I. (2007). *Scientific research in the humanities and social sciences*. Dar Al Hamed Publishing & Distribution, second edition, Amman, Jordan.
- Alhawari, Arwa. (2014). The Effect of Test Length and Ability Distribution Form on the Estimation of A person's Ability According To Rasch Model in Item
- Allen, D. (2016). *The Impact of Shortening a Long Survey on Response Rate and Response Quality*. Brigham Young University.
- Al-Qaisi, A. (2016). The Effects of the Sample Size and the Length of the Test on the Accuracy Estimation of the Item Parameters by Using Non-Parametric Item Response Theory, *Mu'tah Research and Studies, Humanities and Social Sciences Series*, (31)5, 203-246.
- Bani Ata, Z. (2013). Investigating the Accuracy of Estimation of a Three-Parameter Logistic Model of Item Parameter and Individuals' Ability in Light of Test Length and Sample Size: Simulation Study, *Journal of the University of Sharjah for Humanities and Social Sciences*, 11(2), 55-76.

- Breoen, A. & Olivares, A. (2011). Item Response Modeling of Forced Choic Questionnaires, *Educational and Psychological Measurement*, 71(3) 460–502.
- De Ayala, R. (2009). *The Theory and Practice of Item Response Theory*. The Guilford Press, New York.
- Drasgow, F. (1982). Biased test items and differential validity. *Psychological Bulletin*, 92, 526-531.
- E.Bolt, E. & Heide, A. (2013). Reducing questionnaire length did not improve physician response rate: a randomized trial, *Journal of Clinical Epidemiology*, 67(4), 447-481.
- Fontanella, L. Villano, P. & Donato, M. (2015). Attitudes owards Roma people and migrants: a comparison through a Bayesian multidimensional IRT model. *Quality &Quantity*, 49(1), 1-20
- Galesic, M. & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, 73(7), 349–360.
- Golia, S. (2011). The Impact of Questionnaire Size on the Accuracy of the Rasch Measure, *Journal of Applied Science*, 11(4), 707-712.
- Hambleton, R. K. Swaminathan, H. & Rogers, H. (1991). *Fundamentals of item Response Theory*. New York: Sage Publications the International Professional Publishers.
- Hambleton, R. & Swaminathan, H. (1985). *Item response theory principles and applications*. Boston: Kluwer Nijhoff Publishing
- Hambleton, R.K. cook. L.L. (1983). Robutness of item response models and effects of test length and sample size in the precision of ability estimates. NewYork.-InD.J. Weiss(Ed), *New horizons in testing*, pp.31-49.
- Harris, D. (1989). Comparison of 1-, 2-, and 3-parameter IRTmodels. *Educational Measurement: Issues and Practice*. 8, 35-41.

- Herzog, A. R. & Bachman, J. G. (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly*, 45(6), 549–559.
- Howe, K. (1995). *Validity, Bias, and Justice in Educational Testing: The Limits of the consequentialist conception*. Retrieved January 14, 2018, form: <http://www.ed.uiuc.edu/eps/pes-year book //as-does / howe -html>.
- Jansen, L. & Steultjens, M. Holtslag, R. Kwakkel, G. & Dekker, J. (2010). Psychometric properties of questionnaires evaluating health-related quality of life and functional status in polytrauma patients with lower extremity injury, *Journal of trauma Manage Outcomes*, Vol4(7).
- Lord, F. M. (1980). *Application of item response theory to practical testing problems*. NJ: Lawrence Erlbaum Associates Inc.
- Lustina, M. (2004). *A comparison of Andrich rating scale model and Rosts successive intervals model*. [on- Line]: <http://www.lip.utexas.edu/etd/d/2004/Lustinam52242/Lustinam52242.pdf>.
- McMillan, W. (2004). *Classroom assessment: principles and effective instruction*. USA.
- Mond, J. Rodgers, B. Hay, P. Owen, C. & Beumont, P. (2004). Mode of delivery but not questionnaire length, affected response in an epidemiological study of eating disordered behavior. *Journal of Clinical Epidemiolog*, 57(9), 1167-1171.
- Petrillo, J. Cano, S. McLeod, L. & Coon, D. (2015). Using Classical Test Theory, Item Response Theory, and Rasch Measurement Theory to Evaluate Patient Reported Outcome Measures: A Comparison of Worked Examples, *VALUE INHEALTH*, 18, 25 – 34.
- Qandilji, A. (2008). *Scientific research and the use of traditional and electronic sources of information*. Dar Al Masirah, first Edition, Amman, Jordan.

- Reber, R. & Zupanek, N. (2002). Effects of processing fluency on estimates of probability and frequency. In P. Sedlmeier & T. Betsch (Eds.), *Frequent processing and cognition* (pp. 175-188). Oxford, UK: Oxford University Press.
- Ree, M. & Jensen, H. (1983). Effects of sample size on linear equating of item characteristic curve parameters. In D.J. weiss (Ed.), *New horizons in testing*. pp135-146. NewYork.
- Response Theory, *Najah University Journal for Research in Humanities*, 8(29), 44 - 67.
- Rup, A. & Zumbo, B. (2006). Understanding parameter invariance in unidimensional IRT models. *Educational and Psychological Measurement*, 66, 63-84.
- Samejima, F. (1994). Some critical observations of the test information function as a measure of local accuracy in ability estimation. *Psychometrika*, 59 (3), 307-329.
- Samejima, F. (1997). *Graded response model*. In W. J. van der Linden & R. K. Hambleton (Eds). *Handbook of modern item response theory* (pp. 85-100). New York: Springer-Verlag.
- Sass, D. Schmitt, T. & Walker, C. (2004). *An evaluation of BILOG-MG with skewed theta distribution using various estimation procedures: A simulation study*. Poster presented at the National Council on Measurement in Education. San Diego, California.
- Sawalmeh, Y. & Kawasmeh, A. (2000). The impact of diversity in the number of alternatives right and wrong in the multiple Tests of true and false in their psychometric characteristics. *Damascus University Journal* 16 (1), 61 - 88.
- Sawalmeh, Y. (1994). The multiple Tests of true and false: characteristics and the relative number of correct alternatives. *Research Yarmouk: Series Human and Social Sciences* 10 (2), 573 - 595.

- Schofield, H. (1972). *Assessment and Testing: An Introduction*, London: George Allen and Urwin.
- Schumacker, R. (2005). *Test Bias and Differential Item Functioning*. Retrieved January 11, 2009.
- Seong, T. (1990). Sensitivity of Marginal Maximum Likelihood Estimation of item and ability parameter to the characteristics of the prior ability distribution. *Applied Psychological Measurement* ,14, 299-311.
- Stocking, M. (1990). Specifying optimum examinees for item parameter estimation in item response theory. *Psychometrika*. 55(3), 461-475.
- Stone, C. A. (1992). Recovery of Marginal Maximum Likelihood Estimation in Two Parameter Logistic Response Model: An evaluation of MULTILOG. *Applied Psychological Measurement* ,16, 1-16.
- Tay, L. Meade, M. & Smith, M. (2015). An overview and practical guide to IRT measurement equivalence analysis. *Organizational ResearchMethods*,18, 346.
- Thissen, D. & Wainer, H. (1983). Some Standard Errors in Item Response Theory. *Psychometrika*, 47, 397-412.
- Waller, G. Meyer, C. & Ohanian, V. (2001). Psychometric Properties of the Long and Short Versions of the Young Schema Questionnaire: Core Beliefs Among Bulimic and Comparison Women, *Cognitive Therapy and Research*, 25(2), 137–147.
- Warm, T. A. (1978). *A Primer of item Response Theory*. Oklahoma: U. S. Coast Guard Institute 73/69.